HHS Public Access

Author manuscript

Appl Spectrosc. Author manuscript; available in PMC 2017 May 02.

Published in final edited form as:

Appl Spectrosc. 2017 May ; 71(5): 1014–1024. doi:10.1177/0003702816666288.

Direct-on-Filter α-Quartz Estimation in Respirable Coal Mine Dust Using Transmission Fourier Transform Infrared Spectrometry and Partial Least Squares Regression

Arthur L. Miller[1], **Andrew Todd Weakley**[2], **Peter R. Griffiths**[3], **Emanuele G. Cauda**[4], and **Sean Bayman**[1]

[1]National Institute for Occupational Safety and Health (NIOSH), Spokane, WA, USA

[2]IMPROVE Group, Crocker Nuclear Laboratory, University of California Davis, Davis, CA, USA

[3]Griffiths Consulting LLC, Ogden, UT, USA

[4]National Institute for Occupational Safety and Health, Pittsburgh, PA, USA

Abstract

In order to help reduce silicosis in miners, the National Institute for Occupational Health and Safety (NIOSH) is developing field-portable methods for measuring airborne respirable crystalline silica (RCS), specifically the polymorph α-quartz, in mine dusts. In this study we demonstrate the feasibility of end-of-shift measurement of α-quartz using a direct-on-filter (DoF) method to analyze coal mine dust samples deposited onto polyvinyl chloride filters. The DoF method is potentially amenable for on-site analyses, but deviates from the current regulatory determination of RCS for coal mines by eliminating two sample preparation steps: ashing the sampling filter and redepositing the ash prior to quantification by Fourier transform infrared (FT-IR) spectrometry. In this study, the FT-IR spectra of 66 coal dust samples from active mines were used, and the RCS was quantified by using: (1) an ordinary least squares (OLS) calibration approach that utilizes standard silica material as done in the Mine Safety and Health Administration's P7 method; and (2) a partial least squares (PLS) regression approach. Both were capable of accounting for kaolinite, which can confound the IR analysis of silica. The OLS method utilized analytical standards for silica calibration and kaolin correction, resulting in a good linear correlation with P7 results and minimal bias but with the accuracy limited by the presence of kaolinite. The PLS approach also produced predictions well-correlated to the P7 method, as well as better accuracy in RCS prediction, and no bias due to variable kaolinite mass. Besides decreased sensitivity to mineral or substrate confounders, PLS has the advantage that the analyst is not required to correct for the presence of kaolinite or background interferences related to the substrate, making the method potentially viable for automated RCS prediction in the field. This study demonstrated the

Corresponding author: Arthur L. Miller, National Institute for Occupational Safety and Health (NIOSH), 315 E. Montgomery Avenue, Spokane, WA 99207, USA, ALMiller@cdc.gov.

Conflict of Interest

The authors report there are no conflicts of interest. The findings and conclusions in this article are those of the authors and do not necessarily represent the views of the National Institute for Occupational Safety and Health (NIOSH). Mention of any company or product does not constitute endorsement by NIOSH.

efficacy of FT-IR transmission spectrometry for silica determination in coal mine dusts, using both OLS and PLS analyses, when kaolinite was present.

Keywords

Coal mine dust; direct-on-filter Fourier transform infrared; Monte Carlo unimportant variable elimination; partial least squares; silica measurement

Introduction

Sustained occupational exposure to airborne respirable crystalline silica (RCS) has detrimental effects to human health, including decreased lung function as well as increased incidences of pulmonary tuberculosis, lung cancer, and silicosis.[1–4] The U.S. Mine Safety and Health Administration (MSHA) sets daily exposure limits on the most abundant RCS polymorph, α-quartz. Regulatory exposure assessment utilizes two analytical techniques for the analysis of samples collected in mines for RCS, namely mid-infrared (IR) spectrometry and powder X-ray diffraction (XRD). The former approach is mandated for use in U.S. coal mines,[5] while non-coal mine dusts are evaluated using XRD.[6]

Current methods for RCS quantification usually require mailing a sample to an off-site laboratory for analysis. Timely decisions regarding exposure reduction and implementation of control technologies in the workplace are therefore limited. Such time lapses may lead to miners being overexposed to RCS for long periods of time before action is taken, even in environments where the elevated presence of silica has been detected and addressed from a regulatory standpoint.[7] Developing field-based monitoring techniques for RCS is a current research objective of the National Institute for Occupational Safety and Health (NIOSH). The use of mid-IR analysis, particularly Fourier transform infrared (FT-IR) spectrometry, for a direct-on-filter (DoF) determination of RCS has shown promise.[8,9] Prior to recommending the field deployment of IR procedures, the technical limitations associated with RCS quantification using on-site IR (DoF) analysis must be investigated. A primary limitation of a DoF IR-based method for measuring RCS is the presence of siliceous IR confounders in the mine geology.

Coal mine dust typically contains both RCS and kaolinite clay, the IR spectra of which exhibit overlapping absorption features around 800 cm^{-1}. The presence of kaolinite significantly influences the estimated mass of RCS in coal dusts estimated by IR spectrometry and kaolinite correction methods are common to the laboratory determination of RCS.[5,10] The MSHA P7 method minimizes kaolinite interference by adjusting the area of the quartz doublet by referencing the Al–OH vibration which is assumed to correspond only to kaolinite at ~915 cm^{-1}.[11] The P7 protocol also suppresses the organic fraction of coal by ashing the sampling filter and depositing the resulting product onto a new (DM-450 copolymer) filter. Ashing is particularly important for RCS determination in higher-rank coal dusts, such as low-volatile bituminous and anthracite, given the characteristically strong aromatic C–H out-of-plane (OOP) bending bands near 800 cm^{-1}.[12,13] While these steps in the P7 protocol certainly eliminate most common mineral and organic interferences, they

may consequently result in a loss of mass and therefore add bias to the resulting RCS calibration and subsequent estimation of exposure.

Biases associated with sample handling, filter ashing, and ash redeposition are either minimized or reduced by DoF analysis. However, interference from organic material left within the IR sampling volume becomes the primary source of suspected analytical error. Before isolating the impact of organic interferences to the DoF determination of silica, the impact of nonuniform dust deposition, known to occur with the industry-standard tamper-proof cassettes, was ruled out.[14] Specifically, preliminary investigations using a limited set of laboratory-prepared samples concluded that RCS determination using DoF IR analysis for lab-generated coal dusts is not substantially impacted by current sampling equipment and methodologies.[14,15]

The present study extends DoF FT-IR determination of RCS to samples collected from active U.S. coal mines. Two calibration protocols are employed to assess the impact of DoF analysis on RCS determination relative to the MSHA P7 method. The first protocol closely follows the ordinary least squares (OLS) estimation of RCS from the MSHA P7 method by developing calibration factors from analytical silica and kaolinite standards; DoF analysis constitutes the only significant difference between the two methods. The sensitivity of the calibration to organic and mineral interferences is then assessed by predicting RCS in samples acquired from six active coal mines.

The second protocol uses a multivariate partial least squares (PLS) regression to develop a calibration directly from coal mine samples, with an independent laboratory providing the estimated RCS mass to use as standards (y) in the calibration.[16,17] Using field standards for calibration constitutes a secondary calibration and is commonplace in multivariate analysis.[16,17] In the present context, a single-analyte PLS calibration decomposes an N-by-p matrix of FT-IR spectra, [X], by projecting the between-sample variations in IR absorption onto new orthogonal variables (“scores”) which are used to predict RCS mass. Formally, the PLS calibration equations consider every available wavenumber (as opposed to band area or amplitude) when developing regression parameters as follows:

$$[\boldsymbol{X}] = [\boldsymbol{T}]\,[\boldsymbol{P}]^{T} + [\boldsymbol{E}] \quad (1)$$

$$\boldsymbol{y} = [\boldsymbol{T}]\,\boldsymbol{q} + \boldsymbol{f} \quad (2)$$

Here, [$\boldsymbol{X}$] is decomposed into the product of the PLS scores [$\boldsymbol{T}$] and loadings matrix, [$\boldsymbol{P}$]. The scores describe the orthogonal projection of each p-length spectrum (P = 5048 wavenumber in these spectra) onto the PLS components whereas the loadings describing the coefficients of linear combination describing the major between-sample variations in [$\boldsymbol{X}$]. The residual matrix, [$\boldsymbol{E}$], describes all variability in [$\boldsymbol{X}$] that is not modeled by the components and includes additive noise and all other noninterfering absorptions in the spectral matrix. In Eq. 2, we see that the scores matrix— by construction, maximally

correlated to the RCS standards (***y***)— are used to quantify RCS with regression coefficients (***q***) relating the scores to RCS standards with residuals ***f***.

Scores and loadings, often referred to collectively as components, latent variables, or factors,[18,19] are extracted sequentially with the first maximally correlated to the RCS standards and later explaining absorption variability less correlated to the RCS standards.[20] If spectra are numerically preprocessed prior to calibration (e.g., baseline corrected),[21–23] this allows the algorithm to separate absorption variations exclusively associated with RCS mass (e.g., variations in wavenumbers comprising the 800 cm^{-1} band) from variability related to interferences (e.g., polyvinyl chloride [PVC] filter) onto distinct components. Therefore, the first few PLS components (columns in [***T***] and [***P***]) are generally predictive of RCS while the later may correct the regression coefficients used to estimate RCS mass for interferences related to mine geology, slight differences in filter thickness, filter substrate scattering, and so on. Furthermore, since the PLS calibration is developed from field standards, any kaolin correction performed by the independent laboratory is incorporated into the estimated PLS parameters, implying that, unlike the OLS method, another round of kaolinite correction is not needed for the PLS protocol.

For the reasons mentioned above, a PLS regression is expected to produce comparatively better predictions (e.g., lower standard error of prediction, SEP) than the OLS protocol. Furthermore, if spectra are adequately pre-processed to suppress background prior to calibration, a clear and distinct connection to either the analyte (i.e., RCS) or any strongly absorbing interferences (e.g., kaolinite, sampling filter) will be apparent on PLS diagnostic plots (e.g., scores plots, loadings). This has already been shown to be true for RCS determined in non-coal mine samples.[24] Furthermore, the PLS protocol does not require the analyst to directly interact with the FT-IR spectra by performing band integrations, but considers the individual spectral channels (wavenumbers) automatically when performing a calibration. Therefore, the (blind) prediction of field samples is readily accomplished by transforming the regression coefficients (***q***) into the more familiar form as:

$$b = [W]\left([P]^T[W]\right)^{-1} q \quad (3)$$

with the following RCS prediction equation as

$$y_T = [X]_T b + g \quad (4)$$

where [***W***] are the PLS loading weights that are discussed in detail elsewhere.[16,17]

Methodology

Sampling Procedures

For this study, a total of 101 samples of coal mine dust were collected on PVC filters during air quality surveys in six active coal mines. Airborne dust was collected using sampling trains standardized for coal mines (Code of Federal Regulations 30, Part 70.201). The

sampling flow rate was 2.0 L min^{-1} and large (>5 μm) particulate matter was removed using Dorr–Oliver cyclones, which only pass the respirable fraction. Respirable particles were deposited onto pre-weighed 37 mm diameter PVC filters with 5 μm pore size (Zefon, Inc.) mounted within the coal dust sampling cassettes. After collection, all filter samples were weighed to determine the mass of loaded respirable dust. The origin of the samples is summarized in Table 1.

Fourier Transform Infrared Analysis

The filters were removed from the sampling cassettes and mounted in a stainless steel holder for FT-IR analysis. Filters were carefully mounted to ensure that the 6 mm diameter IR beam always probed the center of the filter. Mounting samples in this manner proved adequate for capturing a representative fraction of RCS deposited onto the PVC filter when samples were collected using any one of three common filter cassettes located after a Dorr–Oliver cyclone.[14] Transmission IR spectra of the dust-laden filters were acquired using a Bruker Alpha FT-IR spectrometer. Spectra were collected in transmission mode at 2 cm^{-1} resolution by averaging 40 scans. Interferograms were multiplied using the Blackman–Harris three-term apodization function prior to Fourier transformation. Spectra were saved from 399.5 to 3998.5 cm^{-1}, resulting in each spectrum containing 5084 channels (data points). Each individual spectrum was saved both in its “raw” form and after ratioing it to the single-beam spectrum of the open beam. After samples were scanned they were sent to a laboratory for independent RCS quantification using the MSHA P7 method (RJ Lee Group, Pittsburgh, PA, USA). These independent estimates provided the response variables ($\boldsymbol{y}_{P7}$) needed to develop the PLS calibration and judge the impact of DoF sampling on RCS determination.

RCS Determination by the OLS Protocol

The OLS protocol's capability to determine RCS in coal mines samples is summarized in Figure 1. Raw spectra of laboratory-prepared silica standards and coal mine samples were background corrected using OPUS spectral analysis software (Bruker Optics) prior to manually integrating the α-quartz and kaolinite analytical bands at ~800 and 915 cm^{-1}, respectively.[15] Background correction consisted of ratioing the measured spectrum against the spectrum of a filter blank which was acquired prior to analyzing a batch of dust samples. This procedure eliminated most (but not all) of the absorption spectrum of the filter, since the thickness of each individual filter used for dust samples varied compared to the thickness of the filter blank. Concave rubber band baseline correction was applied (10 iterations) to reduce the impact of baseline on defining the limits of integration for the quartz and kaolinite analytical bands.[25]

The integrated absorbances of the α-quartz doublet and kaolinite bands from standard spectra is then used to determine calibration factors (“OLS calibration;” Fig. 1) as in previous work.[9] Of the orginal 101 mine samples only 66 were used to test the OLS protocol (see “RCS prediction,” Fig. 1). Samples excluded from the analysis had an RCS mass either below the limit of quantification (LOQ) for the OLS method or below the LOQ for the MSHA P7 method.[9] Respirable crystalline silica prediction in the remaining 66 coal mine samples required applying a calibration factor ($CF_{915/800}$) to correct the RCS integrated absorption (A_{RCS}) for kaolinite interference while the second calibration factor

($CF_{800,\mu gQuartz}$) determined the RCS mass in the sample from the corrected absorption measurement.[10,15]

The impact of DoF sampling was evaluated by comparing OLS predictions to an independent laboratory's analysis of the same samples by the MSHA P7 method ($\hat{\boldsymbol{g}} = \hat{\boldsymbol{y}}_{OLS} - \boldsymbol{y}_{P7}$). The standard error of prediction (SEP) and bias for each sample were developed from these comparisons. Bias was calculated for each sample as the percent relative difference between the OLS-predicted and P7 quantification of RCS in each sample, $Bias\,(\%) = 100\left(\frac{\hat{g}}{y_{P7}}\right)$.[26] Average bias and its corresponding 95% confidence limits were estimated to give an understanding of the degree of uncertainty in the sample bias.

RCS Prediction by the PLS Protocol

The PLS protocol is summarized in Figure 2. Preprocessing, including derivative transforming spectra and wavenumber selection, was applied prior to RCS determination to ensure viable RCS predictions and interpretable PLS components. Specifically, the raw 5084-point DoF spectra were transformed to first-derivative spectra using a second-order, 21-channel Savitzky–Golay filter.[27] Derivative filtering has been shown to suppress broad baselines and offsets thereby improving the signal-to-background ratio.[28]

Previous work has documented the benefits of applying wavenumber selection prior to PLS calibration to identify only the most important variables (columns in [$\boldsymbol{X}$]) used for RCS determination in non-coal mine samples.[24,29] In this study, the backward Monte Carlo unimportant variable elimination (BMCUVE) procedure was again tasked with selecting the best variables for RCS determination.[24] Given the limited number of samples available for calibration and prediction (n = 66), data from metal/non-metal mine (MNM) samples containing a known mass of RCS and spectra acquired on the same FT-IR instrument were used for wavenumber selection by BMCUVE (see Weakley et al.[24] for details). This separate data set was used to distance the process of wavenumber selection from PLS calibration, preserving coal mine samples for calibration and RCS prediction. Although MNM mine samples contained no trace of kaolinite absorption, these spectra were still used to remove many of the 5084 wavenumbers that corresponded to excessive background and unreliable silica absorption.

The same 66 sample-pairs used for OLS analysis ([$\boldsymbol{X}$],$\boldsymbol{y}_{P7}$) were next partitioned into two sets, one for model training (calibration) and one for method evaluation (testing). A stratified partitioning of samples was followed to ensure that the distribution and range of RCS mass were comparable in the calibration and testing sets. Specifically, samples from Mines 2, 6, and 6a were assigned as the calibration set (N_C=39) with the remaining samples from Mines 1, 3, 4, and 5 placed in the testing set (N_T=27). Next, two samples from Mine 6a were moved to the testing set to better match the distribution of calibration samples. Spectral matrices were then standardized (“auto-scaled”) to unit variance and response variables ($\boldsymbol{y}_{C,P7}$, $\boldsymbol{y}_{T\,P7}$) mean-centered prior to PLS calibration and RCS prediction.[22,23]

Following transformation, scaling, variable selection, and sample partitioning, PLS regression was carried out using the nonlinear, iterative partial least-squares (NIPALS)

algorithm.[16] The number of PLS components used to predict RCS was determined by fivefold cross-validation and selected according to a minimized root mean squared error of cross-validation (RMSECV).[30] Respirable crystalline silica prediction performance was judged using the same metrics applied to OLS model.

Results and Discussion

The total mass of dust on the 101 samples collected for this study was in the range of 126–2678 μg with the percent silica in the range of 2–30%. The median composition of RCS for the 66 samples above the DoF method LOQ deviated significantly between mine sites, with the shape of the respective distributions and incidence of sample extremes indicating that silica exposure varied markedly across the mine sites. This highlights the need for a DoF method that is robust in relation to fluctuations in RCS mass as well as to variations in mineral and organic content on the sampling filters.

For RCS determination in coal dust samples, both OLS and PLS protocols are potentially hindered by the presence of kaolinite and other organic and inorganic confounders that may be unique to each mine site. In order to visualize these potential interferences, the spectra were grouped according to mine site, averaged, and the most important region for our analysis was plotted (Figure 3). Samples from Mine 6 had the largest average absorbance of kaolinite as seen in Figure 3. Mine 5, on the contrary, contained very little kaolinite but a large absorption band at 878 cm^{-1}. This was presumably related to carbonate absorption but the characteristically strong CO_3^{-2} antisymmetric stretch of carbonate at 1430 cm^{-1} was not identified due to water vapor interference (Figure S4).

At this point, we would like to mention a caveat concerning Figure 3. The IR spectra of the samples collected in this study were predominantly composed of mineral absorption bands (silica, kaolinite) and were remarkably devoid of bands due to organic material. For example, it should be noted that the IR spectra of high-rank coals (anthracites and low-volatile bituminous) show three moderately strong bands between 900 and 700 cm^{-1} that are assigned to aromatic C–H OOP bending modes.[31] The central of the three bands has its maximum absorption at approximately 815 cm^{-1} constituting a potential interference to the α-quartz doublet. The coals investigated in this study were all ranked high-volatile bituminous and yet we see no evidence of these bands. Mine 1 shows a strong absorption at 868 cm^{-1}, which might indicate the presence of aromaticity related to coal particulates; however, bands at ~815 cm^{-1} and 750 cm^{-1} were not resolvable from PVC, thus casting doubt on an aryl C–H OOP bending assignment. Overall, we suspect that the absence of organic constituents on the filters is a result of them failing to pass through the cyclone because their relative particle size is larger than the respirable range. A more detailed investigation into this phenomenon will be the subject of a future study.

A plot of the OLS-predicted ($\hat{\boldsymbol{y}}_{OLS}$) against the measured ($\boldsymbol{y}_{P7}$) RCS values illustrates the accuracy attainable by the DoF protocol (Figure 4). Figure 4 indicates a strong linear correlation between the DoF estimation and the RCS quantities obtained by the independent P7 analysis (R^2 = 0.9477). Average bias and standard error for bias (SE_{Bias}) are small with bias not significantly different from zero (α = 0.05).

Addressing the impact of DoF sampling on RCS prediction was a main task of this study. Figure 5 shows that demarcating the bias plot at 25 μg helps us to visualize why sample predictions show a low average bias for the OLS procedure. Furthermore, Figure 5 suggests that negative sample bias is potentially correlated to the ratio of the quartz doublet area to the kaolinite band area (A_{RCS}/A_{kao}). Samples with bias less than −15% always contained $A_{RCS}/A_{Kao} < 0.91$ with the magnitude of negative bias proportional to A_{RCS}/A_{kao}, e.g., the most negatively biased sample above 25 μg had an $A_{RCS}/A_{Kao} = 0.44$.

Polyvinyl chloride shows strong unresolved absorption centered at ~956 cm^{-1} and, more importantly, a broad shoulder at 919 cm^{-1}.[32,33] Evidence of PVC absorption remaining in the spectra after background correction likely affected the integration of the kaolinite band at 915 cm^{-1}, adding a positive contribution to the estimated band area. The confounding location of these PVC bands therefore helps explain negative prediction bias since a falsely high kaolinite absorption leads to the over-correction of the silica doublet and consequently the under-prediction of RCS mass. Therefore, samples showing a high loading of kaolinite and low mass loading of RCS will likely tend to exhibit negative bias for this OLS calibration.

Prior to developing the PLS calibration, the BMCUVE technique was employed for wavenumber selection using metal/non-metal spectra. The backward Monte Carlo unimportant variable elimination identified an optimal subset of wavenumbers exclusively from the α-quartz doublet region (Figure 6, blue features), closely duplicating results from our previous study.[24] Unlike the case for non-coal samples, kaolinite was readily apparent in the IR spectra of coal dust captured on the filter (Figures 3 and 6a). Thus we anticipated that utilizing only features from the quartz doublet region would lead to somewhat reduced accuracy for RCS prediction, i.e., wavenumber selection provided a strong first step in model optimization but left the decision of whether to incorporate kaolinite confounders to the analyst. This prompted a decision to include 95 additional wavenumbers into the PLS calibration between 889 to 956 cm^{-1} to capture features from the kaolinite and PVC bands (green points in Figure 6a and b).

After wavenumber selection, fivefold cross-validation identified a four-component model as optimal for RCS quantification. Respirable crystalline silica predictions in coal mine samples are shown in Figure 7. Respirable crystalline silica prediction by PLS exhibited an approximately twofold lower SEP (compared to OLS) while using only about one-third of the available samples (27 versus 66) in the calculation. This result was not unexpected given the aforementioned ability of PLS to model RCS and then consign confounder absorption (and noise) to the lesser PLS components, thereby leaving the RCS determination unencumbered.

Partial least squares regression exhibited higher average bias than the OLS model. Confidence intervals indicated no significant difference between the two methods, however. Wider confidence intervals for the PLS method may reflect the increased uncertainty of estimating average bias using only 27 samples versus all 66 samples for the OLS procedure (see Figures 1 and 2 for details). In fact, the twofold drop in SEP is reproduced clearly on bias plots (Figure 8) with PLS prediction errors never exceeding ±35% sample bias

(compared to samples regularly exceeding these margins for the OLS protocol). A trend (dip) in sample bias is evident with most negatively biased samples collected at Mine 1, a batch excluded from the calibration set. Attempts to link bias behavior to gravimetric measurements, A_{RCS}/A_{kao}, or substrate absorption proved unsuccessful. This motivated coloring samples according to their mine affiliation and exploring the FT-IR spectra for mineral and organic absorption unique to Mine 1.

Figure 9 illustrates that the trend in sample bias is traceable to the IR absorption behavior of seven samples, uncovered by a principal component analysis (PCA).[34–36] Projecting each spectrum onto the second and third principal components (PCs) reveals a clustering of Mine 1 samples in the lower left of the subspace. The dispersion of Mine 1 samples on the third PC was largely linked to variations in IR absorption between 860 and 880 cm^{-1}, a range considered unimportant by BMCUVE wavenumber selection. However, the proximity of this region to the kaolinite band (930–885 cm^{-1}) suggests that interference from absorption in the range of 880–860 cm^{-1} was probable (see Figure S6). Specifically, these negatively biased samples (Figure 8) show a moderately strong absorption band at ~868 cm^{-1}, which is evident in the first derivative spectra (Figure 6b) and confirmed in the absorption spectra (Figure S6). Absorption in this region may indicate the presence of the aforementioned aryl C–H OOP bending modes characteristic of higher-rank coals. If correct, an associated absorption band at 814 cm^{-1} is expected to confound the quartz analytical region and therefore positively bias predictions.[13,31] These Mine 1 samples, however, show negative bias and no evidence of other aromatic stretches outside 900–700 cm^{-1}. The cause of the PLS calibration under-predicting RCS in Mine 1 samples therefore remains elusive.

Figure 9 can also be used to assess absorption differences in other samples and connect those differences to prediction bias. For example, one observation from Mine 5 (<25 μg, bias ≈ 31%) appears diametrically opposite to the encircled Mine 1 cluster in Figure 9. This sample also shows a sharp band centered at 878 cm^{-1} (Figure 3). This absorption band directly interfered with the kaolinite stretch at 915 cm^{-1} on the right hand side of the band which likely explains its position on PC_1 and PC_3. It is not possible to link the dispersion (nor relative position) of samples on the principal component plots to definite interference sources without access to detailed mineral, phase, and elemental composition measurements for these mine samples. Regardless, the PCs derived from these spectra are useful in identifying when a sample will show extreme or unique absorption behavior and therefore may fall outside the predictive capabilities of the PLS method. Principal components may therefore prove valuable in RCS field predictions in that samples can be prescreened for interferences prior to prediction by the PLS approach.

Conclusions

Two methods were tested to determine the efficacy of predicting RCS mass in coal dust samples using transmission FT-IR spectroscopy in a DoF application. The first method (based on OLS) is a close analog to the MSHA P7 method and is simple to apply and generally applicable to samples containing confounder-free silica, since it incorporates silica response based on pure silica standards. If kaolinite (the major confounder in coal mine dusts) was present in the field samples, its presence was accounted for by using a secondary

peak from the IR spectrum (~915 cm^{-1}) to enable a direct correction. Employing analytical standards and kaolinite correction to 66 mine dust samples resulted in a linear correlation with P7 results, an acceptable standard error of prediction, and zero average bias according to confidence intervals. A minor limitation of the OLS approach was the sensitivity of the calibration to the proportion of kaolinite present in the field samples, i.e., high kaolinite relative to RCS IR absorption led to the small under-prediction of RCS in coal dust samples, as seen in RCS sample bias.

The second method employed a PLS approach and, while more complex, produced predictions that showed a good correlation to the primary P7 method, twofold greater accuracy in RCS prediction, and no clear correlation between under-prediction and kaolinite mass. Furthermore, the analyst was not required to directly correct the RCS measurement for kaolinite nor background interferences related to the substrate, leaving the method viable for automated RCS prediction in the field. The improved precision and convenience of the PLS method came at the cost of increased sensitivity to mineral or substrate confounders. A complementary PCA allowed us to diagnose the likely source of model bias as absorption concentrated in the region 860–880 cm^{-1}. Overall, the eventual application of a PLS method for the field prediction of RCS will require including many more samples in the calibration set, preferably from mines with a wide range of geological conditions.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors would like to thank Randy Reed and his colleagues at the NIOSH Pittsburgh Mining Research Division for their assistance, as they provided the many coal dust samples from active mines, which made this study possible.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

References

1. Calvert GM, Rice FL, Boiano JM, Sheehy JW, Sanderson WT. Occupational Silica Exposure and Risk of Various Diseases: An Analysis Using Death Certificates from 27 States of the United States. Occup. Environ. Med. 2003; 60(2):122–129. [PubMed: 12554840]
2. Mannetje AT, Steenland K, Attfield M, Boffetta P, Checkoway H, DeKlerk N, Koskela RS. Exposure-Response Analysis and Risk Assessment for Silica and Silicosis Mortality in a Pooled Analysis of Six Cohorts. Occup. Environ. Med. 2002; 59(11):723–728. [PubMed: 12409529]
3. Mazurek JM, Attfield MD. Silicosis Mortality among Young Adults in the United States, 1968–2004. Am. J. Ind. Med. 2008; 51(8):568–578. [PubMed: 18521821]
4. Leung CC, Yu ITS, Chen W. Silicosis. Lancet. 2012; 379(9830):2008–2018. [PubMed: 22534002]
5. MSHA Administration. Infrared Determination of Quartz in Respirable Coal Mine Dust – Method No. MSHA P7. US Dept of Labor-MSHA-Pittsburgh; Safety and Health Technology Center; PA: 2008.
6. Madsen FA, Rose MC, Cee R. Review of Quartz Analytical Methodologies: Present and Future Needs. Appl. Occup. Environ. Hyg. 1995; 10(12):991–1002.

7. Joy GJ. Evaluation of the Approach to Respirable Quartz Exposure Control in U.S. Coal Mines. J. Occup. Environ. Hyg. 2012; 9(2):65–68. [PubMed: 22181563]
8. Miller AL, Drake PL, Murphy NC, Noll JD, Volkwein JC. Evaluating Portable Infrared Spectrometers for Measuring the Silica Content of Coal Dust. J. Environ. Monit. 2012; 14(1):48–55. [PubMed: 22130611]
9. Cauda EG, Miller AL, Drake PL. Promoting Early Exposure Monitoring for Respirable Crystalline Silica: Taking the Laboratory to the Mine Site. J. Occup. Environ. Hyg. 2016; 13(3):D39–45. [PubMed: 26558490]
10. Ainsworth S. Infrared Analysis of Respirable Coal Mine Dust for Quartz: Thirty-Five Years. Journal of ASTM International. 2005; 2(4):13.
11. Frost RL. The Structure of the Kaolinite Minerals; a FT-Raman Study. Clay Minerals. 1997; 32(1): 65–77.
12. Painter PC, Coleman MM, Jenkins RG, Whang PW, Walker PL Jr. Fourier Transform Infrared Study of Mineral Matter in Coal. A Novel Method for Quantitative Mineralogical Analysis. 1978; 57(6):337–344.
13. Fuller MP, Hamadeh IM, Griffiths PR, Lowenhaupt DE. Diffuse Reflectance Infrared Spectrometry of Powdered Coals. Fuel. 1982; 61(6):529–536.
14. Miller AL, Drake PL, Murphy NC, Cauda EG, LeBouf RF, Markevicius G. Deposition Uniformity of Coal Dust on Filters and Its Effect on the Accuracy of FTIR Analyses for Silica. Aerosol Sci. Technol. 2013; 47(7):724–733. [PubMed: 26719603]
15. Miller AL, Drake PL, Murphy NC, Noll JD, Volkwein JC. Evaluating Portable Infrared Spectrometers for Measuring the Silica Content of Coal Dust. J. Environ. Monit. 2012; 14(1):48–55. [PubMed: 22130611]
16. Geladi P, Kowalski BR. Partial Least-Squares Regression: A Tutorial. Anal. Chim. Acta. 1986; 185(0):1–17.
17. Wold S, Sjöström M, Eriksson L. PLS-Regression: A Basic Tool of Chemometrics. Chemometrics and Intelligent Laboratory Systems. 2001; 58(2):109–130.
18. Abdi H. Partial Least Squares Regression and Projection on Latent Structure Regression (PLS Regression). Wiley Interdisciplinary Reviews: Computational Statistics. 2010; 2(1):97–106.
19. Malinowski, ER. Factor Analysis in Chemistry. Wiley; Hoboken, NJ: 2002.
20. Geladi P, MacDougall D, Martens H. Linearization and Scatter-Correction for Near-Infrared Reflectance Spectra of Meat. Appl. Spectrosc. 1985; 39(3):491–500.
21. Xu L, Zhou YP, Tang LJ, Wu HL, Jiang JH, Shen GL, Yu RQ. Ensemble Preprocessing of Near-Infrared (NIR) Spectra for Multivariate Calibration. Anal. Chim. Acta. 2008; 616(2):138–143. [PubMed: 18482596]
22. Bocklitz T, Walter A, Hartmann K, Rosch P, Popp J. How to Pre-Process Raman Spectra for Reliable and Stable Models? Anal. Chim. Acta. 2011; 704(1–2):47–56. [PubMed: 21907020]
23. Berg RA, Hoefsloot HC, Westerhuis JA, Smilde AK, Werf MJ. Centering, Scaling, and Transformations: Improving the Biological Information Content of Metabolomics Data. BMC Genomics. 2006; 7(1):1. [PubMed: 16403227]
24. Weakley AT, Miller AL, Griffiths PR, Bayman S. Quantifying Silica in Filter-Deposited Mine Dusts Using Infrared Spectra and Partial Least Squares Regression. Anal. Bioanal. Chem. 2014; 406(19):4715–4724. [PubMed: 24830397]
25. Wartewig, S. IR and Raman Spectroscopy: Fundamental Processing. Wiley; Hoboken, NJ: 2003.
26. Kennedy, ER., Fischbach, TJ., Song, R., Eller, PM., Shulman, SA. Guidelines for Air Sampling and Analytical Method Development and Evaluation. National Institute for Occupational Safety and Health; Cincinnati, OH: 1994.
27. Savitzky A, Golay MJE. Smoothing and Differentiation of Data by Simplified Least Squares Procedures. Anal. Chem. 1964; 36(8):1627–1639.
28. Rinnan Å, van den Berg F, Engelsen SB. Review of the Most Common Pre-Processing Techniques for Near-Infrared Spectra. TrAC, Trends Anal. Chem. 2009; 28(10):1201–1222.
29. Mehmood T, Liland KH, Snipen L, Sæbø S. A Review of Variable Selection Methods in Partial Least Squares Regression. Chemometrics and Intelligent Laboratory Systems. 2012; 118(0):62–69.

30. Arlot S, Celisse A. A Survey of Cross-Validation Procedures for Model Selection. Statistics Surveys. 2010; 4:40–79.
31. Solomon PR, Carangelo RM. FT-IR Analysis of Coal. Fuel. 1988; 67(7):949–959.
32. Tabb DL, Koenig JL. Fourier Transform Infrared Study of Plasticized and Unplasticized Poly(Vinyl Chloride). Macromolecules. 1975; 8(6):929–934.
33. Krimm S. Infrared Spectroscopy and Polymer Structure. Pure Appl. Chem. 1968; 16(2):19.
34. Cowe IA, McNicol JW. The Use of Principal Components in the Analysis of Near-Infrared Spectra. Appl. Spectrosc. 1985; 39(2):257–266.
35. Nieuwoudt HH, Prior BA, Pretorius IS, Manley M, Bauer FF. Principal Component Analysis Applied to Fourier Transform Infrared Spectroscopy for the Design of Calibration Sets for Glycerol Prediction Models in Wine and for the Detection and Classification of Outlier Samples. J. Agric. Food. Chem. 2004; 52(12):3726–3735. [PubMed: 15186089]
36. Abdi H, Williams LJ. Principal Component Analysis. Wiley Interdisciplinary Reviews: Computational Statistics. 2010; 2(4):433–459.
37. Wold S, Esbensen K, Geladi P. Principal Component Analysis. Chemometrics and Intelligent Laboratory Systems. 1987; 2(1–3):37–52.

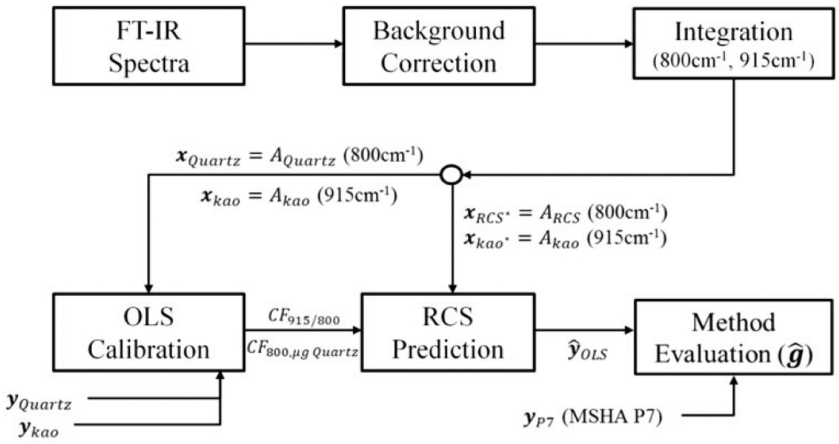

Figure 1.

Summary of the OLS protocol. The FT-IR spectra of filter-deposited quartz standards and coal mine samples are first background corrected to remove PVC absorption and residual baseline. Band integration, calibration, and RCS prediction faithfully follow the MSHA P7 method wherein quartz and kaolinite absorption are estimated by integrating the α-quartz doublet (~800 cm^{-1}) and Al–OH stretch (~915 cm^{-1}) in laboratory standard spectra ($\boldsymbol{A_{Quartz}}$,$\boldsymbol{A_{kao}}$) followed by estimation in field samples ($\boldsymbol{A_{RCS}}$,$\boldsymbol{A_{kao}}$). Calibration factors (*CF*) are then developed from quartz standards, and RCS predicted in the coal mine samples. The method and, by extension, impact of DoF sampling are evaluated by comparing OLS ($\boldsymbol{\hat{y}_{OLS}}$) to independent MSHA P7 predictions ($\boldsymbol{y_{P7}}$).

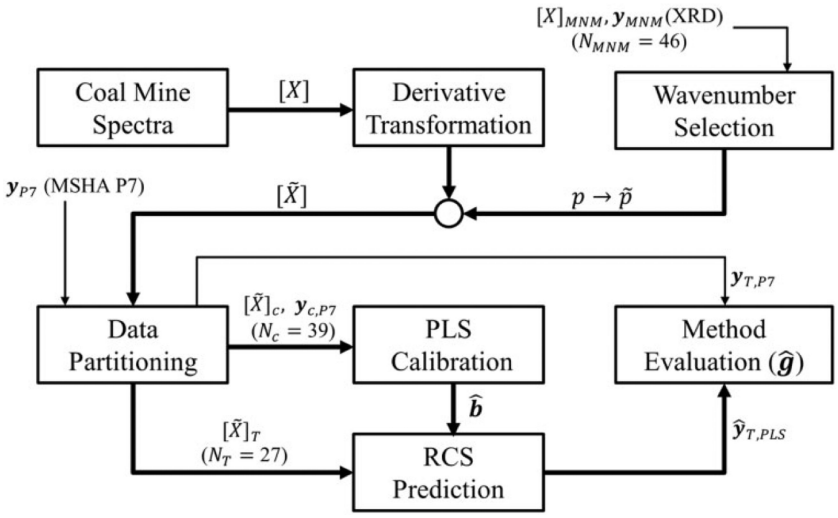

Figure 2.

Summary of PLS protocol for RCS determination in coal mine samples. The 66 coal mine spectra are transformed to first-derivative spectra and a fraction ($\tilde{\boldsymbol{p}}$) of the total 5084 data points ($\boldsymbol{p}$) are selected using data from metal/non-metal (MNM) mine dust samples from Weakley et al.[24] Preprocessed spectra, [$\tilde{\boldsymbol{X}}$], and P7 estimates, $\boldsymbol{y}_{P7}$, are partitioned into calibration and testing sets (subscripted "c" and "*T*," respectively). PLS calibration develops regression coefficients ($\hat{\boldsymbol{b}}$) to predict RCS in test samples with method evaluation then qualifying prediction performance. The tilde (~) denotes that wavenumber selection has been performed.

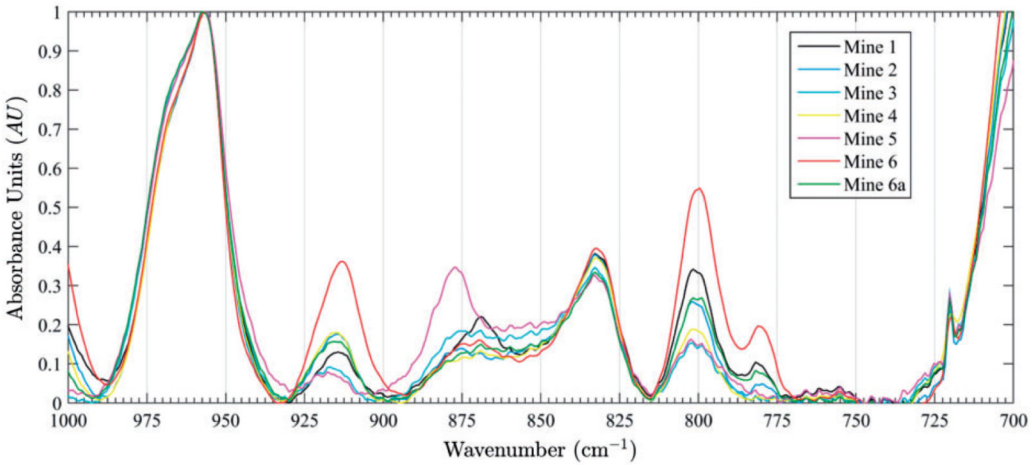

Figure 3.

Average baseline-corrected spectra from each mine site, normalized to the PVC band at 956 cm^{-1}. Silica and kaolinite vibrations are evident between 817 and 767 cm^{-1} and 930 and 885 cm^{-1}, respectively. Polyvinyl chloride vibrations are identifiable at 956 cm^{-1} and 832 cm^{-1} with detailed information about the character of substrate absorption confined to supporting information.

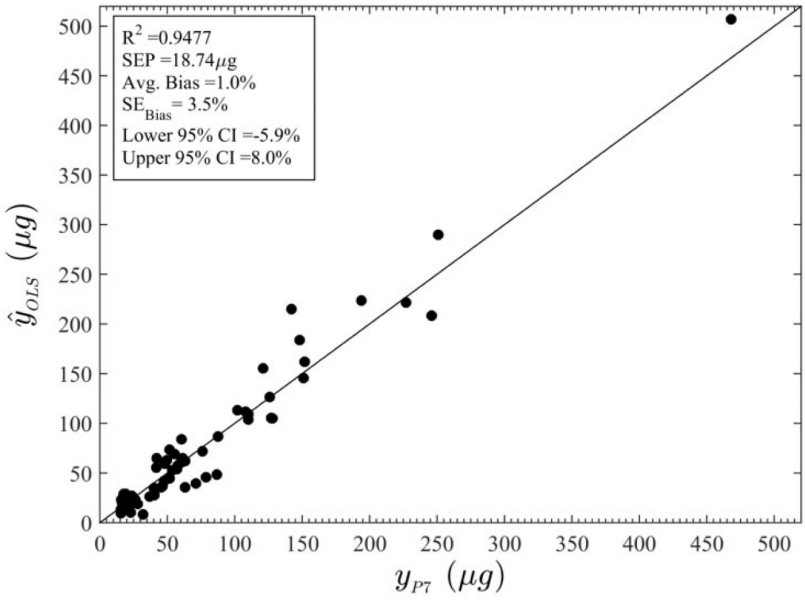

Figure 4.

Ordinary least-squares-predicted RCS content plotted with respect to the corresponding MSHA P7 estimates. The standard error of prediction (SEP), average bias, standard errors for bias (SE_{Bias}), and 95% confidence intervals (CI) for the average bias are shown in the figure insert.

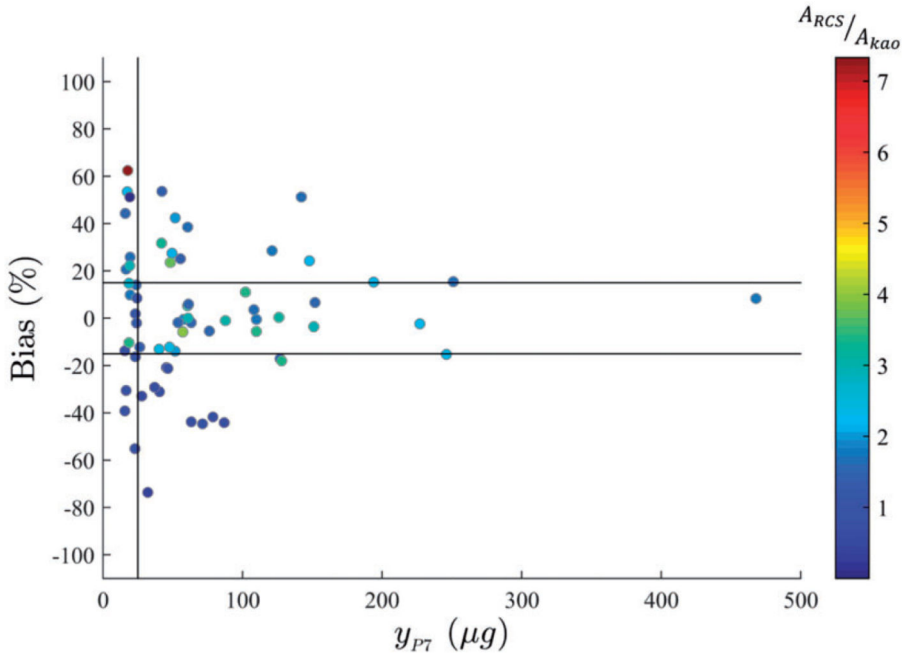

Figure 5.

Predicted sample bias from OLS regression pseudo-colored according to $\boldsymbol{A_{RCS}/A_{kao}}$. Samples colored dark blue indicate a low $\boldsymbol{A_{RCS}/A_{kao}}$ while dark red indicates the converse.

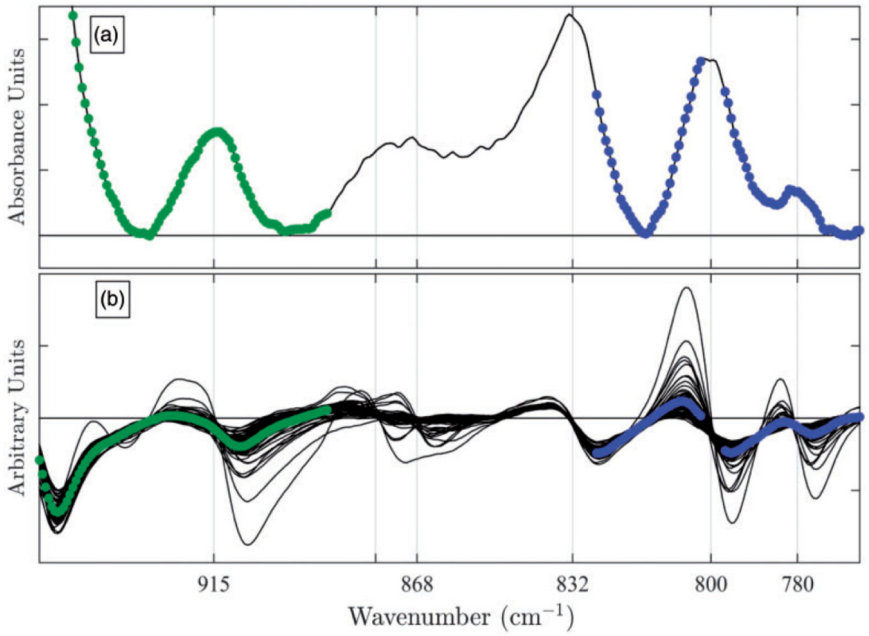

Figure 6.

Analytical region selected by BMCUVE and used for PLS calibration and prediction. An average absorption spectrum (a) and all first derivative spectra (b) were calculated and plotted for the 66 mine samples, respectively. Blue wavenumbers indicate those chosen by the BMCUVE feature selection using non-coal mine samples (80 variables). Green wavenumbers (95 variables) were chosen manually to cover the substrate and kaolinite interference regions. Important band centers are indicated by vertical lines in the derivative spectra, with some spectra showing distinctive behavior at 868 and 878 cm^{-1} (unlabeled).

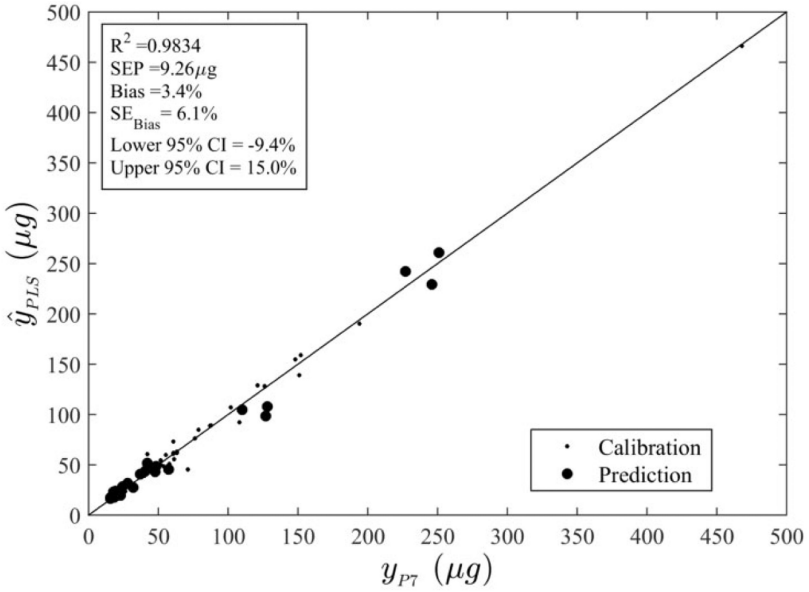

Figure 7.

Respirable crystalline silica predictions by the PLS protocol plotted with respect to the corresponding MSHA P7 estimates. The SEP, average bias, standard error for bias (SE_{Bias}), and 95% confidence limits for bias are shown. The number of PLS components and variables used are also indicated.

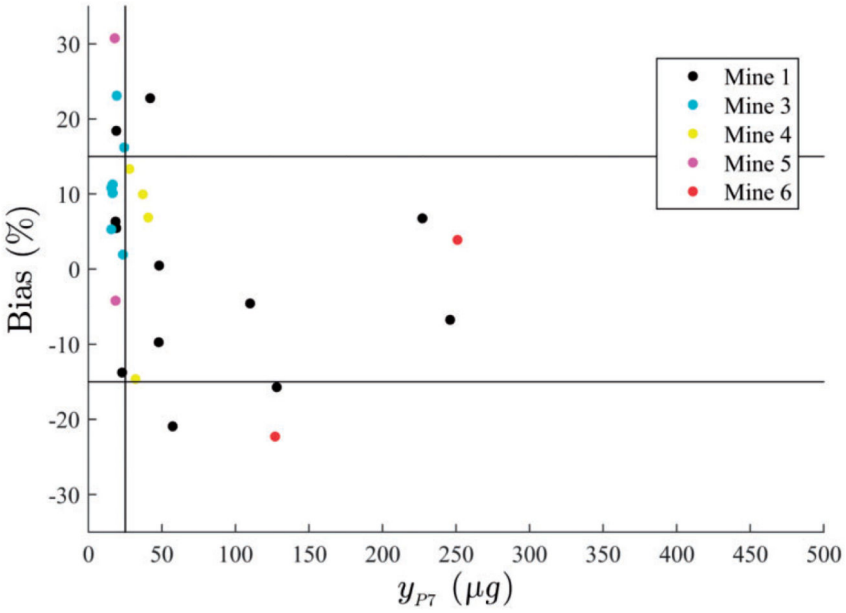

Figure 8.
Predicted sample bias from the PLS protocol.

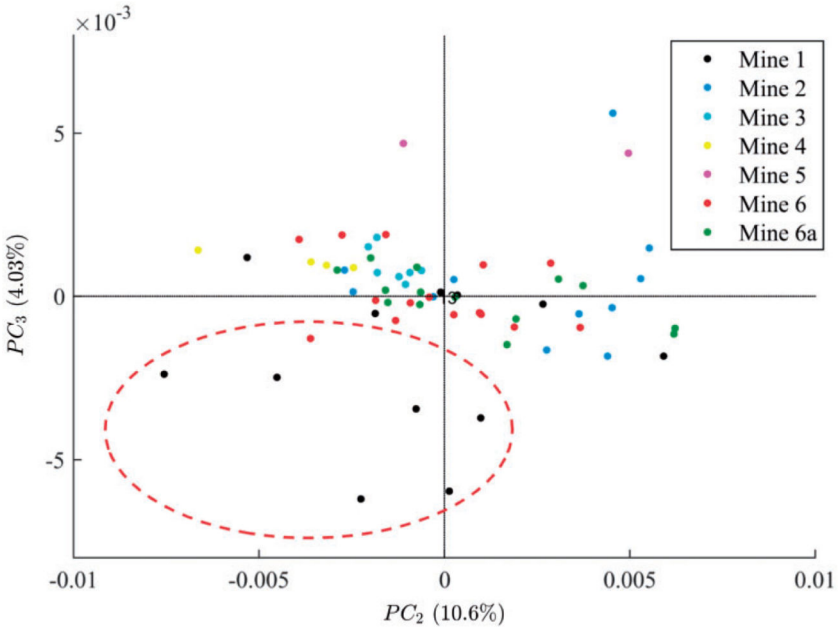

Figure 9.

The result of projecting derivative spectra onto their second and third PCs. The PCs were calculated using all wavenumbers in the range of 765–956 cm^{-1}. The legend indicates the sample's affiliation to a given mine. Parentheses denote proportion of variance explained by each component (see Wold et al.[37] for interpretation of PC score plots).

Table 1

Geographic, operating, and geologic characteristics of the six U.S. coal mines sampled in this study. Samples acquired from Mines 6 and 6a were collected from the same site on different occasions. The number of total samples (#) and samples above the OLS method limit of quantification (LOQ) are displayed as well.

Batch	Operation	County	State	Seam	Rank	Samples (#, > LOQ)
Mine 1	Underground	Raleigh	WV	Upper Eagle Coal	High-Volatile Bituminous B	(17, 12)
Mine 2	Underground	Harlan	KY	Upper and Lower Harlan	High-Volatile Bituminous B	(16, 11)
Mine 3	Underground	Wise	VA	Upper Parsons	High-Volatile Bituminous B	(14, 7)
Mine 4	Underground	Wayne	WV	Coalburg	High-Volatile Bituminous B	(8, 4)
Mine 5	Underground	Randolph	IL	Herrin No. 6	High-Volatile Bituminous C	(16, 2)
Mine 6	Surface	Kanawha	WV	Kittanning	High-Volatile Bituminous B	(15, 15)
Mine 6a	Surface	Kanawha	WV	Kittanning	High-Volatile Bituminous B	(15, 15)